

## **4.8. Spatial interpolation**

### **4.8.1. Mean annual maximum (or “Index flood”) grids**

As explained in Section 4.6.1, mean annual maximum values were used as the site-specific scaling factor to generate precipitation frequency estimates from regional growth factors (RGFs). The station mean annual maximum values were spatially interpolated to produce mean annual maximum, or “index flood”, grids using technology developed by Oregon State University’s Spatial Climate Analysis Service (SCAS). SCAS has developed PRISM (Parameter-elevation Regressions on Independent Slopes Model), a hybrid statistical-geographic approach to mapping climate data (Daly and Neilson, 1992; Daly et al., 1994; Daly et al., 1997). PRISM spatially interpolated the HDSC-calculated mean annual maximum values by using a naturally strong relationship with mean annual precipitation.

SCAS adapted PRISM to use their existing mean annual precipitation grids (USDA-NRCS, 1998), transformed using the square-root, as the predictor grid for interpolating mean annual maximum precipitation to a uniformly spaced grid. Mean annual precipitation was used as the predictor because it is based on a large data set, accounts for spatial variation of climatic information and is consistent with methods used in previous projects, including NOAA Atlas 2. PRISM uses a unique regression function for each target grid cell and has the ability to account for: user knowledge, the distance of an observing station to the target cell, if the station is in a cluster of stations grouped together, the difference between station and target cell mean annual precipitation, topographic facet, and coastal proximity. Other parameters include radius of influence, minimum number of stations on a facet, and total number of stations required for the regression to estimate the mean annual maximum precipitation at a given grid cell. PRISM cross-validation statistics were computed where each observing station was deleted from the data set one at a time and a prediction made in its absence. Results indicated that any overall bias was less than 2 percent and mean standard error was about 10 percent for this Atlas. Appendix A.4 provides additional information regarding the details of the work done by SCAS for HDSC.

Table 4.8.1 lists the mean annual maximum (a.k.a. “index flood”) grids, one for each duration of the project, that were interpolated by PRISM. The resulting high-resolution (30-second, or about 0.5 mile x 0.5 mile) mean annual maximum grids then served as the basis for deriving precipitation frequency estimates at different return periods using a unique HDSC-developed spatial interpolation procedure, the Cascade, Residual Add-Back (CRAB) derivation procedure (described in detail in Section 4.8.2).

Deviations may occur between the observed point mean annual maximum values in the HDSC database and the resulting grid cell value due to spatial interpolating and smoothing techniques employed by PRISM. The HDSC database consists of station metadata (longitude, latitude, period of record, etc.), the mean annual maximum values and precipitation frequency estimates for each station. These deviations occur because PRISM produces interpolated values that mitigate differences between the observed point estimates and surrounding stations with similar climate, mean annual precipitation, elevation, aspect, distance from large water bodies and rain-shadow influences. See Appendix A.4 for more details.

Table 4.8.1. Mean annual maximum grids interpolated by PRISM.

<b>Duration</b>
60-minute
120-minute
3-hour
6-hour
12-hour
24-hour
48-hour
4-day
7-day
10-day
20-day
30-day
45-day
60-day
<b>Total</b>
<b>14</b>

#### **4.8.2. Derivation of precipitation frequency grids**

The Cascade, Residual Add-Back (CRAB) derivation procedure is a unique spatial interpolation technique, developed by HDSC, to convert mean annual maximum grids into grids of precipitation frequency estimates (see Figure 4.8.1). It accommodates spatial smoothing and interpolating across “region” boundaries to eliminate potential discontinuities due to different RGFs as a result of the regional L-moment analysis.

The CRAB philosophy was first used in the derivation of several of the National Climatic Data Center’s new Climate Atlas maps (Plantico et al., 2000). The CRAB process, as the term *cascade* implies, uses the previously derived grid to derive the next grid in a cascading fashion. The technique derives grids along the frequency dimension with quantile estimates for different durations being separately interpolated. Hence, duration-dependent spatial patterns evolve independently of other durations.

The derivation process utilizes the strong, linear relationship between a particular duration and frequency, the *predictor* estimates, and the next higher frequency of the same duration. Figure 4.8.2 shows the relationship between the *predictor* precipitation frequency estimates, 50-year 24-hour in this example, and the subsequent precipitation frequency estimates, 100-year 24-hour. The R-squared value here of 0.9986 is very close to 1.0 which was common throughout all of the regressions. Since this is calculated using all stations in the project area, the slope of this relationship (1.1345) can be thought of as an average domain-wide regional growth factor (RGF). Regional differences are then accounted for using residuals, as explained below.

A summary of the complete CRAB derivation procedure is illustrated in Figure 4.8.1 and can be summarized in a series of steps. In this description, the term *predictor* refers to the previous grid upon which the subsequent grid is based.

Figure 4.8.1. Flowchart of the cascade residual add-back (CRAB) grid derivation procedure beginning with the mean annual maximum grid of the x-duration and deriving the 2-year x-duration grid as an example.

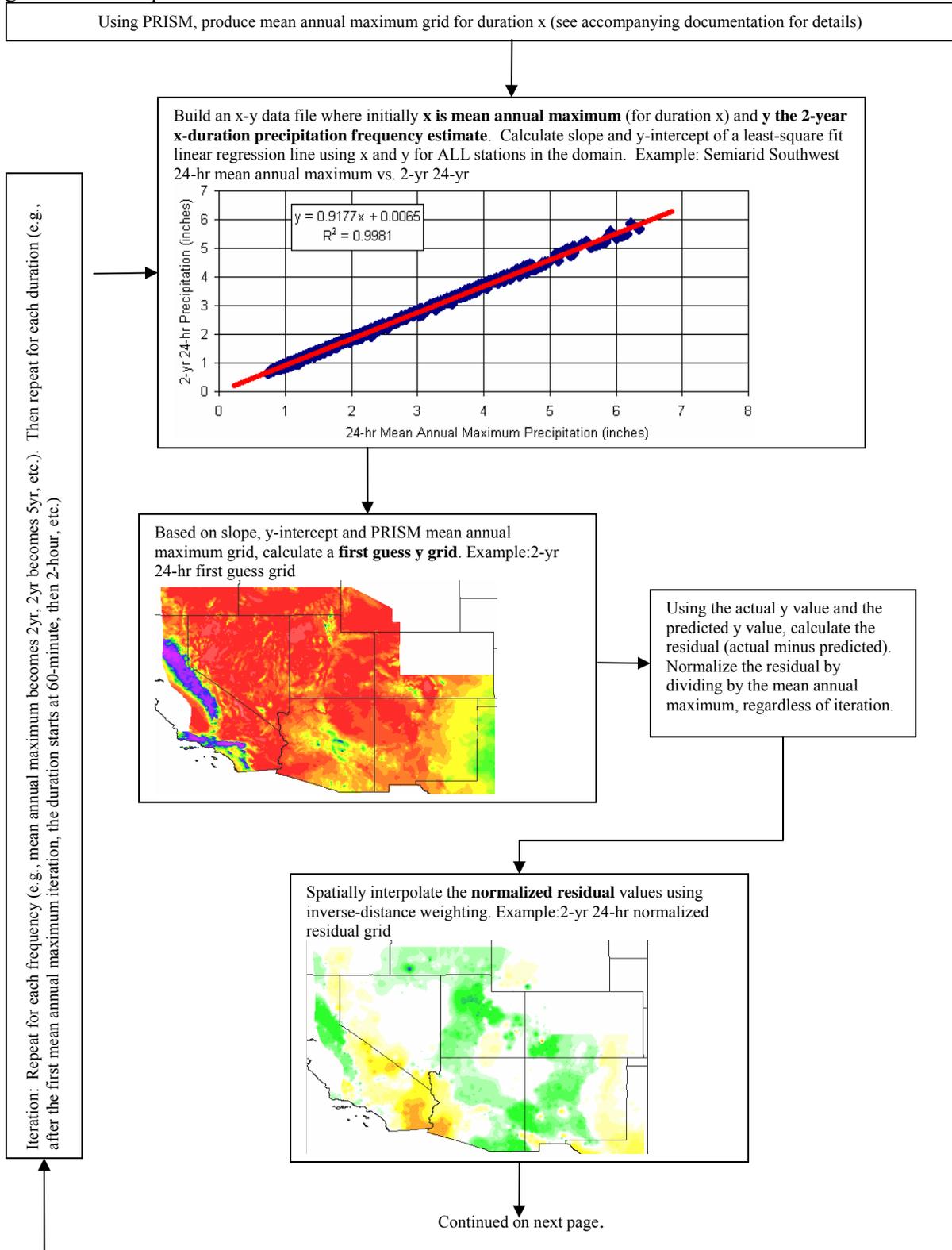
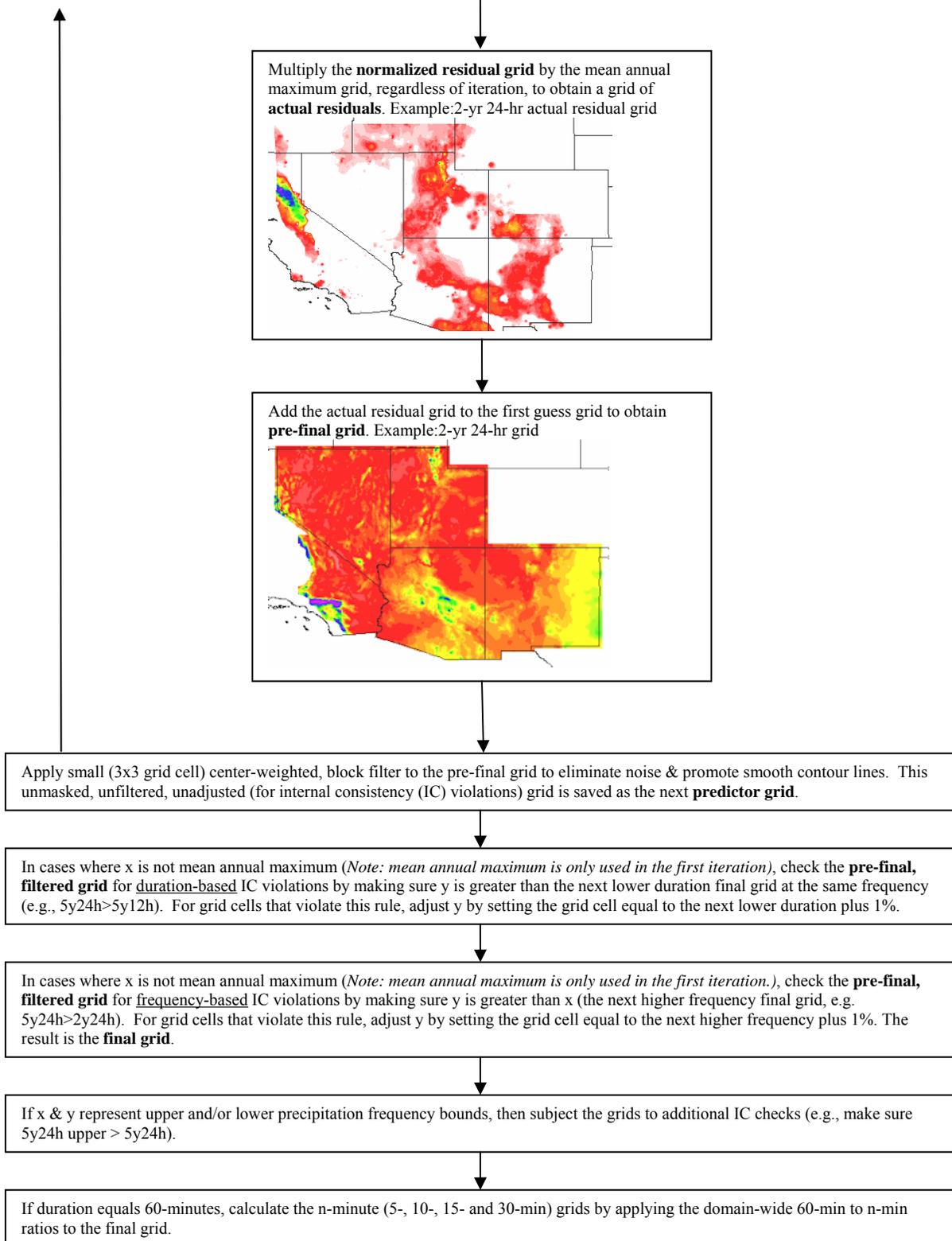


Figure 4.8.1. cont'd



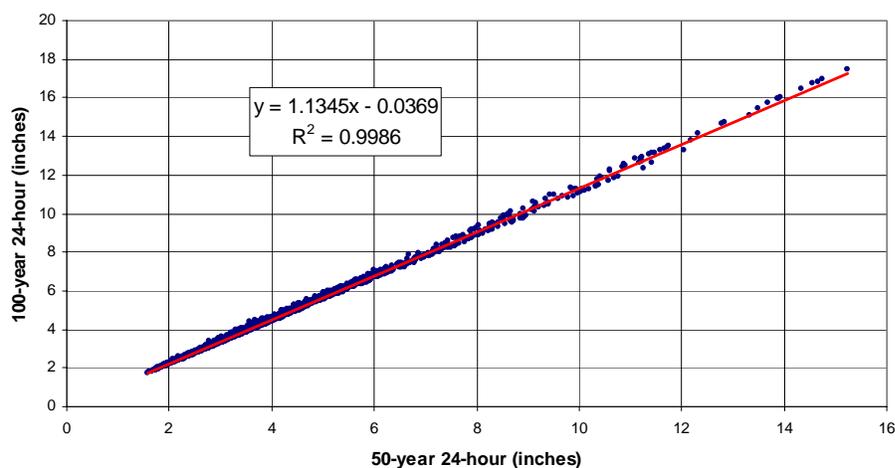


Figure 4.8.2. A scatter plot of 100-year 24-hour vs. 50-year 24-hour precipitation frequency estimates and the linear regression line from NOAA Atlas 14 Volume 1.

**Step 1: Development of regression.** The cascade begins with the mean annual maximum grid derived by SCAS using PRISM for a given duration as the initial *predictor grid* (e.g., 24-hour mean annual maximum) and the 2-year frequency as the subsequent grid (e.g., 2-year 24-hour). All precipitation frequency estimates in the HDSC database were adjusted to make the regression consistent with the PRISM mean annual maximum grids. The adjustment, which is based on the difference between the mean annual maximum PRISM grid cell value and the point mean annual maximum as computed from observed data, was a station-unique factor applied to the precipitation frequency data. The factor is independent of frequency. For instance, a station has an observed mean annual maximum 60-minute value (from the database) of 0.82 inches, but the PRISM grid cell at this station has a value of 0.861 inches. This results in an adjustment factor of 1.05 which is applied to each of the 60-minute precipitation frequencies (2-years through 1,000-years) before constructing the regression equation. These adjusted precipitation frequency estimates are known as the *actual estimates*. In most cases, this adjustment was  $\pm 5\%$  (See Appendix A.4 for more details). A global (all-region) relationship for each duration/frequency pair was developed at the beginning of each iteration based on station precipitation frequency estimates, adjusted for smoothing, at all stations.

To develop the global relationship, an x-y data file was built where initially x was the mean annual maximum for a given duration and y the 2-year precipitation frequency estimate for that duration for each observing station. The slope and y-intercept of a least-square fit linear regression line using x and y for all stations in the domain was calculated. In each region the slope of the line is equal to the regional growth factor (RGF).

**Step 2: Development of first guess grids.** The global linear regression relationship was then applied, using a GIS, to the *predictor grid* (e.g., 24-hour mean annual maximum) to establish a *first guess grid* (e.g., 2-year 24-hour) that was not necessarily equivalent to the *actual estimates* which are based on the unique RGF for each region.

**Step 3: Development of spatially interpolated residual grids.** To account for the regional differences, residuals (*actual estimates* minus *predicted estimates*) at each station were calculated. Here, *predicted estimates* (e.g., 2-year 24-hour) were those derived in the *first guess grid*. The

residuals were normalized by the mean annual maximum to facilitate the interpolation of residuals to ungauged locations.

The normalized residuals at each station were then spatially interpolated to a grid using the Geographic Resources Analysis Support System or GRASS © (GRASS, 2002) GIS inverse-distance-weighting (IDW) algorithm to produce a *normalized residual grid*. The IDW method assumes the value at an unsampled point can be estimated as a weighted average of points within a certain distance or from a given number of *m* closest points; CRAB used the 12 closest points (i.e., *m* = 12). Weights are inversely proportional to the power of the distance which at an unsampled point *r* = (x,y) is:

$$F(r) = \frac{\sum_{i=1}^m z(r_i) / |r - r_i|^p}{\sum_{j=1}^m 1 / |r - r_j|^p} \quad (\text{E.8, Neteler and Mitasova, 2002})$$

where

- $F(r)$  = interpolated precipitation at unsampled grid cell
- $z$  = precipitation at sample point
- $m$  = 12
- $p$  = 2
- $r_{i,j}$  = location of sample point
- $r$  = location of unsampled grid cell.

IDW was used because by definition it is an exact interpolator and remains faithful to the *normalized residuals* at stations; this is important so that when the *normalized residuals* are converted back to *actual residuals* they are equal to the original *actual residual* at each station. Since there is a great deal of spatial autocorrelation of the *normalized residuals*, i.e. the *normalized residuals* tend to be spatially consistent within the regions, IDW was an adequate and appropriate interpolation scheme (see embedded map of normalized residuals in Figure 4.8.1).

The *normalized residual grid* was de-normalized by multiplying it by the original spatially interpolated mean annual maximum grid to obtain a spatially interpolated grid of *actual residuals* for the entire project area. Figure 4.8.3 shows the relationship between the 100-year 24-hour *actual residuals* and the 24-hour mean annual maximum estimates. Each linear cluster shown on this scatter plot represents stations within the same region that have varying 100-year 24-hour precipitation depths.

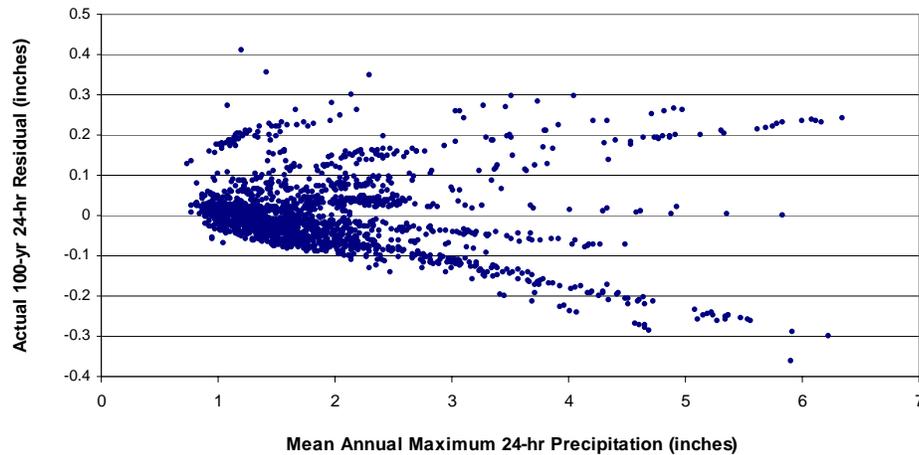


Figure 4.8.3. The relationship between the 100-year 24-hour *actual residuals* and the mean annual maximum precipitation from NOAA Atlas 14 Volume 1.

**Step 4: Development of *pre-final grids*.** The spatially interpolated grid of *actual residuals* was added to the *first guess grid* to create a spatially interpolated *pre-final grid* (e.g., 2-year 24-hour). To remove extraneous noise in the *pre-final grid* and encourage smooth contour lines, a 3x3 grid cell block average filter was applied. To prevent error propagation potentially introduced in the internal consistency adjustment steps (described in Step 5), the *pre-final grid* was archived and became the *predictor grid* for the next precipitation frequency grid derivation. For example, the *pre-final* 2-year 24-hour grid was used as the predictor for the 5-year 24-hour grid rather than the *final* 2-year 24-hour grid to remain faithful to the data and allow patterns to develop without any differences that may be introduced by adjustments and filters.

**Step 5: Internal consistency check.** To ensure internal consistency in the *pre-final grid* cell values, duration-based and frequency-based internal consistency checks were conducted. Frequency-based internal consistency violations (e.g., 100-year < 50-year) were very rare and when they did exist, they were small violations relative to the precipitation frequency estimates involved. Duration-based internal consistency violations (e.g., 24-hour > 12-hour) were more common, particularly between 120-minute and 3-hour, but again were small violations relative to the magnitude of precipitation frequency estimates. To mitigate internal consistency violations, the longer duration or lower frequency grid cell value was adjusted by multiplying the shorter duration or lower frequency grid cell value by 1.01 to provide a 1% difference between the grid cells. One percent was chosen over a fixed factor to allow the difference to change according to the grid cell magnitudes while at the same time providing a minimal, but sufficient, adjustment without changing otherwise compliant data in the process. The duration-based check and adjustment was conducted first, resulting in a new *pre-final grid*, which was then subjected to the frequency-based check and adjustment. The resulting grid became the *final grid* for the particular frequency and duration (e.g., 2-year 24-hour).

**Development of n-minute grids.** Durations shorter than 60-minute (i.e., n-minute precipitation frequency estimates) were calculated using linear scaling factors applied to *final grids* of spatially interpolated 60-minute precipitation frequency estimates. Because there were so few n-minute stations in the project area, global ratios of n-minute to 60-minute estimates were averaged over the entire study area (Section 4.1.1). Using these ratios (listed again in Table 4.8.2), the *final* 60-minute

*grids* were multiplied by the appropriate ratio to compute the appropriate n-minute grid. These ratios were used for all frequencies as well as both the n-minute upper- and lower- confidence limit grids.

Table 4.8.2. NOAA Atlas 14 Volume 1 n-minute ratios: 5-, 10-, 15- and 30-minute to 60-minute.

Duration	5-minute	10-minute	15-minute	30-minute
Ratio	0.318	0.484	0.600	0.808

**Validation.** The initial draft mean annual maximum, “index flood”, grids for this Atlas, as well as the CRAB-derived 100-year 24-hour and 100-year 60-minute precipitation frequency grids were subjected to a peer-review (Appendix A.6). After considering and resolving all reviewer comments, final mean annual maximum grids were created by PRISM and the CRAB procedure re-run.

In addition, jackknife cross-validation allowed further, objective evaluation and validation of the precipitation frequency grids. The jackknife cross-validation exercise entailed running the CRAB procedure with a station in the dataset, storing the target grid cell value (at the station), then running CRAB without the station and comparing the target grid cell values. It was cost prohibitive to re-create the PRISM mean annual maximum grids for each cross validation iteration. For this reason, the cross-validation results reflect the accuracy of the CRAB procedure based on the same mean annual maximum grids. The comparison was used to test the robustness and accuracy of the CRAB interpolation. A perfect validation would result in equal values – with and without the station. 100-year 60-minute results, which required the most interpolation to unengaged locations because of the low number of hourly stations, indicated that the CRAB process performed well (Figure 4.8.4).

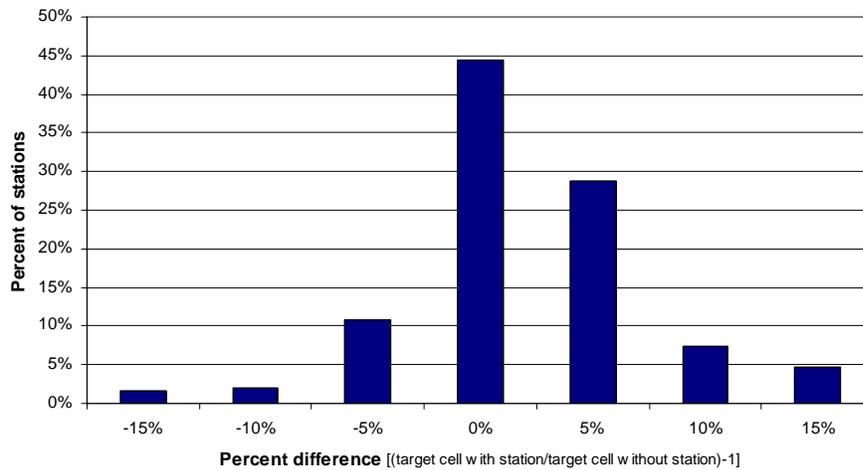


Figure 4.8.4. NOAA Atlas 14 Volume 1 100-year 60-minute jackknife cross-validation results.

## Derivation of upper/lower limit precipitation frequency grids

The upper and lower limit precipitation frequency grids were also derived using the CRAB procedure. Testing suggested that the best method by which to derive the upper/lower limit grids was to use the preceding upper (or lower) grid as the *predictor grid* and *normalizing grid* for the upper/lower limit grid being derived, as opposed to using the corresponding mean precipitation frequency grid. Although the upper (lower) limit precipitation frequency estimates were slightly less stable, they still exhibited strong linear relationships with the previous (predictor) grid. The appropriate (i.e., same duration) mean annual maximum grid (PRISM-produced “index flood”) was used as the initial *predictor grid* for the 2-year upper and lower limit precipitation frequency estimate grids. Figure 4.8.5 shows a scatter plot of the 24-hour mean values versus the 2-year 24-hour upper limit precipitation frequency estimates.

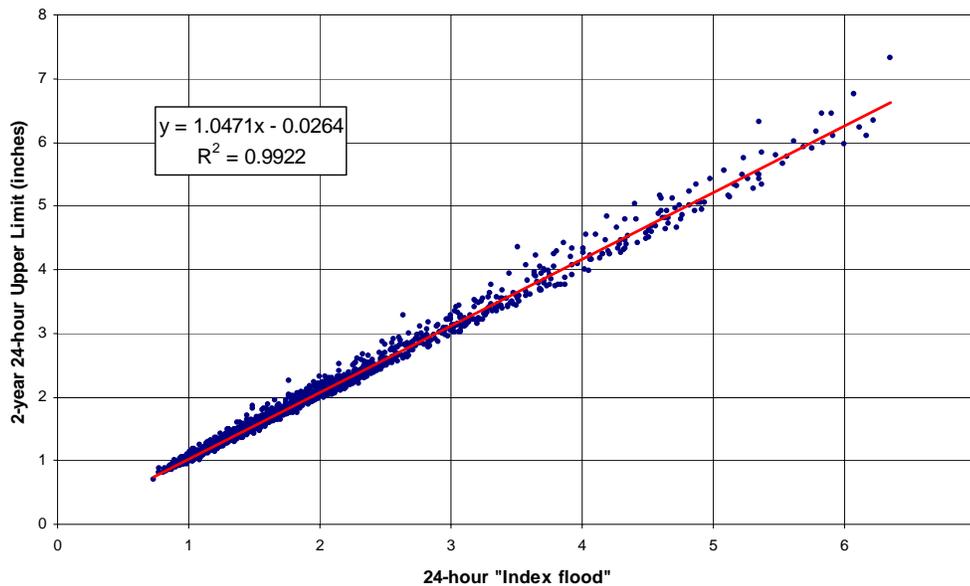


Figure 4.8.5. Scatter plot of the 24-hour mean precipitation frequency estimates vs. the 2-year 24-hour upper limit showing a correlation of 0.9922 in NOAA Atlas 14 Volume 1.

Similar to the precipitation frequency estimate grids, the upper and lower limit grids were evaluated and adjusted for internal consistency. Although very rare, duration-based adjustments were made to ensure the upper (lower) limit grid cell values were larger (smaller) than the mean values. In the event of a violation (e.g., 100-year 60-minute < 100-year 60-minute lower limit) the upper (lower) limit grid was adjusted up (down) by 1% of the mean grid. Like the precipitation grids, frequency-based or duration-based adjustments were made when needed. To mitigate any internal consistency violations, the longer duration or lower frequency grid cell value was adjusted by multiplying the shorter duration or lower frequency grid cell value by 1.01 to provide a 1% difference between the grid cells.

### 4.8.3. Pseudo data

Since each duration was computed independently, it was possible for inconsistencies from duration to duration at a given location to occur. In the spatial interpolation, this was a particular concern at hourly-only and daily-only station locations. However, such inconsistencies were rare.

At hourly-only station locations, inconsistencies could occur because calculated 60-minute through 48-hour estimates anchored the interpolation while 4-day through 60-day estimates at those locations were computed during the spatial interpolation process that was based on estimates at nearby daily stations. During the evaluation phase of the grids, HDSC discovered 6 cases where inconsistencies in the precipitation frequency estimates from 48-hour to 4-day were observed. Each of the cases was resolved after reviewing the observed data and the behavior of nearby stations. In some cases it was clear that the 48-hour data derived from the hourly observations was less reliable than that derived from the daily observations. In these cases, the 48-hour point estimates were removed and instead estimated by spatial interpolation. In the remaining cases, the patterns were not inconsistent with possible climatologies in the area and thus were retained.

Likewise, there were 21 cases where inconsistencies arose at daily-only station locations because calculated 24-hour through 60-day estimates anchored the interpolation while 60-minute through 12-hour estimates at those locations were computed during the spatial interpolation process that was based on estimates at nearby hourly stations. In these 21 cases, the  $\leq 12$ -hour interpolated precipitation frequency estimates were considerably lower and inconsistent with the surrounding calculated  $\geq 24$ -hour precipitation frequency estimates. This caused unreasonable changes in the precipitation frequency estimates from 12-hours to 24-hours at those locations.

These cases were objectively identified by using grids that indicated the difference between the 100-year 12-hour and 100-year 24-hour precipitation frequency estimates. By using these grids, spatial artifacts were differentiated from climatologically-driven patterns. In general, if the difference between the 100-year 12-hour and 100-year 24-hour grid cell value was  $\geq 1.40$ ", the daily-only stations in that area were scrutinized. 21 locations with such inconsistencies were identified and verified for data accuracy. These locations were primarily in desert locations, particularly in southwestern Arizona.

Table 4.8.3. Hourly pseudo stations used in the preparation of NOAA Atlas 14 Volume 1.

Station ID	Station Name	State
02-2434	DATELAND WHITEWING RCH	AZ
02-4702	KOFA MINE	AZ
02-5627	MOHAWK	AZ
02-8396	TACNA 3 NE	AZ
02-9211	WELLTON	AZ
02-9652	YUMA CITRUS STATION	AZ
02-9654	YUMA PROVING GROUND	AZ
02-9656	YUMA QUARTERMASTER DEPOT	AZ
02-9657	YUMA VALLEY	AZ
02-9662	YUMA WB CITY	AZ
04-2319	DEATH VALLEY	CA
04-2506	DOYLE 4 SSE	CA
04-3489	GOLD ROCK RANCH	CA
04-3710	HAIWEE	CA
04-9671	WILDROSE R S	CA
26-0150	AMARGOSA FARMS GAREY	NV
26-0150	AMARGOSA FARMS GAREY	NV
26-6691	RED ROCK CANYON ST PK	NV
29-1138	BOSQUE DEL APACHE	NM
42-2607	ESKDALE PSEUDO	UT
42-5733	MOAB RADIO	UT

So-called pseudo data were used to mitigate the inconsistencies at these 21 locations. Table 4.8.3 lists the hourly pseudo stations generated for this Atlas. The creation of pseudo hourly precipitation

frequency estimates was similar to the approach used to alleviate 12-hour to 24-hour inconsistencies at co-located stations (Section 4.6.2). The pseudo precipitation frequency estimates were generated by applying a ratio of x-hour estimates to 24-hour estimates that was spatially interpolated using GRASS's © inverse-distance-weighting algorithm (GRASS, 2002) based on only co-located daily/hourly stations. The ratio at each co-located station was calculated using the station's 24-hour precipitation frequency estimate to its x-hour precipitation frequency estimate. The interpolated ratio was then applied to the daily-only 24-hour precipitation frequency estimates to generate the pseudo hourly data at that station location. The mitigation provided a smoother, more meteorologically-sound transition from hourly to daily precipitation frequency estimates.

Tests showed that creating pseudo hourly data for daily-only stations that did not exhibit a large difference from 12-hour to 24-hour resulted in nearly identical precipitation frequency estimates before and after the inclusion of pseudo data. Pseudo data were not added to stations that did not need it or at ungauged locations. Locations where an inconsistency between 12-hour and 24-hour estimates could not be expressly proved were assumed accurate based on climate and not mitigated. Pseudo data were used only where deemed absolutely necessary to produce consistent results.

#### **4.8.4. Derivation of isohyets of precipitation frequency estimates**

Isohyetal (contour) GIS files were created from the grids of partial duration series based precipitation estimates for users with geographical information systems (GISs). The isohyets are provided as Environmental Systems Research Institute, Inc. line shapefiles (ESRI, 2003). The isohyets were created by contouring the grid files with GRASS's © r.contour command (GRASS, 2002). The resulting files were when exported as shapefiles with GRASS's © v.out.shapefile command (GRASS, 2002). In order to keep the isohyets and grids consistent, no line generalization or smoothing was conducted. The precision and resolution of the grids were sufficiently high to result in smooth contour lines.

The choice of contour intervals was determined by an algorithm which used the maximum, minimum and range of grid cell values. The number of individual contour intervals was constrained between 10 and 30; however some of the n-minute grids did not exhibit the range necessary to meet the 10 interval threshold and therefore have fewer than 10. All of the intervals are evenly divisible by 0.10 inches – the finest interval. A script that computed the appropriate contour intervals and shapefiles, also generated Federal Geographic Data Committee compliant metadata for the shapefiles and a “fact” file. The HTML-formatted fact file provides details of the shapefile and also includes a list of the contour intervals. To simplify the downloading of the isohyetal shapefiles from the Precipitation Frequency Data Server (PFDS), all of the shapefile components (\*.shp, \*.dbf, and \*.shx, \*.prj), metadata and fact file were compiled and compressed into a single archive file containing many files (\*.tar). For projection, resolution and other details of the shapefiles, please refer to the metadata and/or fact file.

#### **4.8.5. Creation of color cartographic maps**

The isohyetal shapefiles were used to create color cartographic maps of the partial duration series-based precipitation frequency grids. The maps were created using Environmental Systems Research Institute, ArcGIS© 8.3 software, in particular ArcMap© (ESRI, 2003). Although in appearance the cartographic maps look to be comprised of polygons, enclosed two-dimensional cells, they are not. Instead, color shading of the grids combined with the line shapefiles provides the clean look of polygons. The cartographic maps are provided in an Adobe Portable Document format (PDF) format for easy viewing and printing. The scale of the maps is 1:2,000,000 when printed in their native size, 15.5” x 21.5” (same size as the NOAA Atlas 2 maps), however the maps can be printed at any size.